NASA TECHNICAL REPORT

NASA TR R-244

NASA TR R-244

LINEAR PREDICTION ON A FINITE PAST OF A MULTIVARIATE STATIONARY PROCESS

by Ray G. Langebartel

Goddard Space Flight Center
Greenbelt, Md.

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION • WASHINGTON, D. C. • JULY 1966

NASA TR R-244

LINEAR PREDICTION ON A FINITE PAST OF A MULTIVARIATE STATIONARY PROCESS

By Ray G. Langebartel

Goddard Space Flight Center
Greenbelt, Md.

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

For sale by the Clearinghouse for Federal Scientific and Technical Information
Springfield, Virginia 22151 – Price $2.00

ABSTRACT

An overview is given of multivariate, wide-sense, stationary, stochastic process, linear prediction theory with emphasis on the mean square error of prediction based on a finite past. Examples of several different types of processes are examined with specific formulas for the prediction errors obtained in some cases. A discussion is given concerning the effect on the prediction error of basing the prediction on a subprocess of the given process instead of on the entire original process.

CONTENTS

LINEAR PREDICTION ON A FINITE PAST OF A MULTIVARIATE STATIONARY PROCESS

by
Ray G. Langebartel*
Goddard Space Flight Center

INTRODUCTION

Linear prediction theory for one-dimensional stationary stochastic processes ("time series") has been studied extensively (for the essential aspects see, e.g., Reference 1). Less attention has been paid to multivariate processes but a body of theory is now at hand (References 2 to 5). Prediction on a finite past in a multivariate process with discrete parameters is not dwelt on in these works, however. In the present paper this aspect of the theory is considered with special emphasis on the mean square error of prediction. In certain quite special multivariate processes fairly explicit results are obtained, but the problem for more general processes is difficult.

BASIC CONCEPTS

Let X_m, where $m = 0, \pm 1$, and $\pm 2, \cdots$, represent a multivariate discrete stochastic process of dimension M:

$$X_m = \begin{bmatrix} x_m^1 \\ x_m^2 \\ \vdots \\ x_m^M \end{bmatrix} ,$$

a column matrix of M rows. The x_m^k may be complex numbers, in which case we designate the conjugate of x_m^k by $\overline{x_m^k}$ and the conjugate of the matrix X_m by $\overline{X_m}$, the matrix of the conjugates. The transpose of a matrix will be designated by T in this manner: X_m^T. The expected value of a multivariate process is taken as the matrix of the expected values of the elements:

$$E\{X_m\} = \begin{bmatrix} E\{x_m^1\} \\ E\{x_m^2\} \\ \vdots \\ E\{x_m^M\} \end{bmatrix} .$$

*GSFC consultant from the University of Illinois.

The covariance matrix (or correlation matrix) $E\{X_m \overline{X_n^T}\}$ is defined accordingly:

$$E\{X_m \overline{X_n^T}\} = \begin{bmatrix} E\{x_m^1 \overline{x_n^1}\} & E\{x_m^1 \overline{x_n^2}\} & \cdots & E\{x_m^1 \overline{x_n^M}\} \\ E\{x_m^2 \overline{x_n^1}\} & E\{x_m^2 \overline{x_n^2}\} & \cdots & E\{x_m^2 \overline{x_n^M}\} \\ \vdots & & \ddots & \vdots \\ E\{x_m^M \overline{x_n^1}\} & \cdots & \cdots & E\{x_m^M \overline{x_n^M}\} \end{bmatrix}$$

$$\equiv \begin{bmatrix} r_{m,n}^{11} & r_{m,n}^{12} & \cdots & r_{m,n}^{1M} \\ r_{m,n}^{21} & r_{m,n}^{22} & \cdots & r_{m,n}^{2M} \\ \vdots & \vdots & \ddots & \vdots \\ r_{m,n}^{M1} & r_{m,n}^{M2} & & r_{m,n}^{MM} \end{bmatrix} .$$

In the event that each of the elements of this matrix, i.e., the various covariances of the x_m^k processes, depends on m and n only through their difference the multivariate process is called *stationary in the wide sense* (or *weakly stationary*) and we write

$$E\{X_m \overline{X_n^T}\} \equiv R_{(m-n)} \equiv R_{m-n} .$$

It is a straightforward computation to show that the covariance matrix of a wide-sense stationary multivariate process satisfies the relation

$$\overline{R_{(m)}^T} = R_{(-m)} . \tag{1}$$

It is often the case in practice that the elements of the matrices X_n and R_n are of different magnitudes and it is desirable to scale them to some norm. The elements of X_n can be scaled by premultiplying X_n by a diagonal constant matrix A,

$$A = \begin{bmatrix} k_1 & 0 & 0 & \cdots & 0 \\ 0 & k_2 & 0 & \cdots & 0 \\ 0 & 0 & k_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & & k_M \end{bmatrix} ,$$

to give

$$\hat{X}_n = AX_n = \begin{bmatrix} k_1 x_n^1 \\ k_2 x_n^2 \\ \vdots \\ k_M x_n^M \end{bmatrix} .$$

For the scaled covariance matrix $\hat{R}_m$ we then have

$$\hat{R}_m \equiv E\{\hat{X}_{m+n}\overline{\hat{X}_n^T}\} = E\{AX_{m+n}\overline{(AX_n)^T}\} = E\{AX_{m+n}\overline{X_n^T}\,\overline{A^T}\} = AE\{X_{m+n}\overline{X_n^T}\}\overline{A^T} = A R_m \bar{A} .$$

It is customary to choose the elements of A so that the elements of R_0 down the main diagonal are all equal to unity. In the case of real matrices this means that A is

$$\begin{bmatrix} \frac{1}{\sqrt{R_0^{11}}} & 0 & 0 \;\cdots & 0 \\ 0 & \frac{1}{\sqrt{R_0^{22}}} & 0 \;\cdots & 0 \\ 0 & 0 & \frac{1}{\sqrt{R_0^{33}}}\cdots & 0 \\ \vdots & \vdots & \vdots \;\ddots & \vdots \\ 0 & 0 & 0 \;\cdots & \frac{1}{\sqrt{R_0^{MM}}} \end{bmatrix}$$

where $R_0^{ij} = E\{x_n^i\, x_n^j\}$, the elements of R_0. For $M = 2$ the scaled covariance matrix is, when written out,

$$\hat{R}_m = \begin{bmatrix} \frac{R_m^{11}}{R_0^{11}} & \frac{R_m^{12}}{\sqrt{R_0^{11}\, R_0^{22}}} \\ \frac{R_m^{21}}{\sqrt{R_0^{11}\, R_0^{22}}} & \frac{R_m^{22}}{R_0^{22}} \end{bmatrix} = \begin{bmatrix} \frac{E\{x_{m+n}^1\, x_n^1\}}{R_0^{11}} & \frac{E\{x_{m+n}^1\, x_n^2\}}{\sqrt{R_0^{11} R_0^{22}}} \\ \frac{E\{x_{m+n}^2\, x_n^1\}}{\sqrt{R_0^{11} R_0^{22}}} & \frac{E\{x_{m+n}^2\, x_n^2\}}{R_0^{22}} \end{bmatrix} .$$

One way to investigate the nature of a process is to treat its Fourier representation. The wide-sense stationary process X_m has the *spectral distribution function* (square matrix) $F(\lambda)$ representing the covariance matrix,

$$\hat{R}_{(m)} = \int_{-1/2}^{1/2} e^{2\pi i\lambda m}\, dF(\lambda) , \tag{2}$$

with a corresponding representation of the process itself,

$$X_m = \int_{-1/2}^{1/2} e^{2\pi i m\lambda}\, dY(\lambda) , \tag{3}$$

where $Y(\lambda)$ is a column matrix (Reference 1, p. 596; Reference 6, p. 45). There must, of course, be some relation between $F(\lambda)$ and $Y(\lambda)$ for these representations to hold. In terms of the spectral density function $F'(\lambda)$, this relation is

$$E\left\{dY(\lambda)\overline{dY(\mu)^T}\right\} = F'(\lambda)\,\delta(\lambda-\mu)\,d\lambda\,d\mu \tag{4}$$

where $\delta(\lambda)$ is the Dirac delta function. The $Y(\lambda)$ process is spoken of as having "orthogonal increments." The reason for the form of Equation 4 is indicated by the following heuristic calculation

$$\begin{aligned}
R_{(m)} = E\left\{X_{m+n}\overline{X_n^T}\right\} &= E\left\{\int_{-1/2}^{1/2} e^{2\pi i(m+n)\lambda}\,dY(\lambda)\int_{-1/2}^{1/2} e^{-2\pi i n\mu}\,\overline{dY(\mu)^T}\right\} \\
&= E\left\{\int_{-1/2}^{1/2}\int_{-1/2}^{1/2} e^{2\pi i(m+n)\lambda-2\pi i n\mu}\,dY(\lambda)\,\overline{dY(\mu)^T}\right\} \\
&= \int_{-1/2}^{1/2}\int_{-1/2}^{1/2} e^{2\pi i(m+n)\lambda-2\pi i n\mu}\,E\left\{dY(\lambda)\,\overline{dY(\mu)^T}\right\} \\
&= \int_{-1/2}^{1/2}\int_{-1/2}^{1/2} e^{2\pi i(m+n)\lambda-2\pi i n\mu}\,F'(\lambda)\,\delta(\lambda-\mu)\,d\lambda\,d\mu \\
&= \int_{-1/2}^{1/2} e^{2\pi i m\lambda}\,F'(\lambda)\,d\lambda = \int_{-1/2}^{1/2} e^{2\pi i m\lambda}\,dF(\lambda)\ ,
\end{aligned}$$

which is independent of n and thereby confirms wide-sense stationarity. In a particular example $F(\lambda)$ may well have jumps so that $F'(\lambda)$ would not exist in the classical sense (although it would in the Schwartz distribution theory sense) and this is the reason that in much of the literature these Fourier representations are presented in the Stieltjes form.

If the process is real the representations of Equations 2 and 3 can be put in a real form. Let $dY = dY_1 + i dY_2$, so that

$$X_m = \int_{-1/2}^{1/2}\left[(\cos 2\pi m\lambda)\,dY_1 - (\sin 2\pi m\lambda)\,dY_2\right] + i\left[(\cos 2\pi m\lambda)\,dY_2 + (\sin 2\pi m\lambda)\,dY_1\right]\ .$$

This is evidently real if dY_2 is an odd function of λ and dY_1 is even (i.e., if $Y_2'(\lambda)$ is even and $Y_1'(\lambda)$ is odd), in which case the real part of the integrand is an even function of λ and only the half interval (0, 1/2) is required:

$$X_m = 2 \int_0^{1/2} \left[(\cos 2\pi m\lambda)\, dY_1 - (\sin 2\pi m\lambda)\, dY_2 \right] .$$

Let

$$2dY_1 \equiv dU$$

and

$$-\, 2dY_2 \equiv dV$$

and, by substitution, obtain the representation for real values of X_m

$$X_m = \int_0^{1/2} (\cos 2\pi m\lambda)\, dU(\lambda) + (\sin 2\pi m\lambda)\, dV(\lambda) . \tag{5}$$

A corresponding treatment on Equations 2 and 4 leads to the companion formulas

$$R_{(m)} = \int_0^{1/2} (\cos 2\pi\lambda m)\, dG(\lambda) , \tag{6}$$

$$\left.\begin{aligned} E\{dU(\lambda)\, dU^T_{(\mu)}\} &= E\{dV(\lambda)\, dV^T_{(\mu)}\} = G'(\lambda)\, \delta(\lambda - \mu)\, d\lambda\, d\mu , \\ E\{dU(\lambda)\, dV^T_{(\mu)}\} &= E\{dV(\lambda)\, dU^T_{(\mu)}\} = 0 . \end{aligned}\right\} \tag{7}$$

It is natural to introduce orthogonality terminology in view of the relations satisfied by dY, dU, and dV. In particular, we say that X and Y (column matrices with random variables as elements) are *orthogonal* if

$$E\{\overline{XY^T}\} = 0 . \tag{8}$$

Furthermore, X and Y are *orthonormal* if in addition to Equation 8 they satisfy

$$E\{\overline{XX^T}\} = E\{\overline{YY^T}\} = I \tag{9}$$

where I is the identity matrix. An expression of the form $\sum_{n=1}^{N} A_n X_n$ where $X_1, X_2, \cdots X_N$ are multivariates of dimension M and $A_1, \cdots A_N$ are $M \times M$ matrices of constants is called a *linear* *combination* of the X_n. The collection of all linear combinations of $X_1, \cdots X_N$ forms the *linear manifold* generated by $X_1, \cdots X_N$. The *orthogonal projection* of Y on a linear manifold is Y_p if Y_p belongs to the manifold and $Y - Y_p$ is orthogonal to every element of the manifold.

LINEAR PREDICTION

Suppose in a wide-sense, stationary, multivariate, stochastic process X_m we wish to predict X_{N+1} given the values of $X_0, X_1, X_2, \cdots X_N$, and that this predicted X_{N+1} can be represented as a linear combination of $X_0, \cdots X_N$. One method of attack is to introduce Ξ_{N+1}, which is to be the unit normal of the manifold spanned by $X_0, \cdots X_N$, and to represent it as a linear combination of $X_0, \cdots X_N, X_{N+1}$:

$$\Xi_{N+1} = \sum_{r=0}^{N+1} B_r X_r .$$

Then

$$0 = E\left\{X_s \overline{\Xi_{N+1}^T}\right\} = E\left\{X_s \sum_{r=0}^{N+1} \overline{X_r^T}\, \overline{B_r^T}\right\}$$

$$= E\left\{\sum_{r=0}^{N+1} X_s \overline{X_r^T}\, \overline{B_r^T}\right\}$$

$$= \sum_{r=0}^{N+1} E\left\{X_s \overline{X_r^T}\right\} \overline{B_r^T}$$

$$= \sum_{r=0}^{N+1} R_{s-r} \overline{B_r^T} \qquad \text{for} \quad 0 \leq s \leq N ,$$

and

$$I = E\left\{\Xi_{N+1} \overline{\Xi_{N+1}^T}\right\} = E\left\{\sum_{r=0}^{N+1} B_r X_r \overline{\Xi_{N+1}^T}\right\}$$

$$= \sum_{r=0}^{N+1} B_r E\left\{X_r \overline{\Xi_{N+1}^T}\right\}$$

$$= B_{N+1} E\left\{X_{N+1} \overline{\Xi_{N+1}^T}\right\}$$

$$= B_{N+1} \sum_{r=0}^{N+1} R_{N+1-r} \overline{B_r^T} .$$

These two results can be combined to give the matrix equation for $C_r \equiv \overline{B_r^T} B_{N+1}$, provided B_{N+1} is nonsingular,

$$\sum_{r=0}^{N+1} R_{s-r}\, C_r \;=\; \delta_{s,N+1}\, I\,, \qquad \text{for} \quad 0 \leqq s \leqq N+1\,, \tag{10}$$

where $\delta_{s,r}$ is the Kronecker delta. But

$$X_{N+1} \;=\; B_{N+1}^{-1}\; \Xi_{N+1} - \sum_{r=0}^{N} B_{N+1}^{-1}\, B_r\, X_r \;\equiv\; B_{N+1}^{-1}\, \Xi_{N+1} + \Psi\,,$$

so if we use Ψ as the predicted value for X_{N+1} then the mean square error is

$$\begin{aligned}
E\left\{(X_{N+1} - \Psi)\, \overline{(X_{N+1} - \Psi)^T}\right\} &= E\left\{B_{N+1}^{-1}\, \Xi_{N+1}\, \overline{(B_{N+1}^{-1}\, \Xi_{N+1})^T}\right\} \\
&= E\left\{B_{N+1}^{-1}\, \Xi_{N+1}\, \overline{\Xi_{N+1}^T}\; \overline{B_{N+1}^{-1T}}\right\} \\
&= B_{N+1}^{-1}\; E\left\{\Xi_{N+1}\, \overline{\Xi_{N+1}^T}\right\}\, \overline{B_{N+1}^{-1T}} \\
&= B_{N+1}^{-1}\, \overline{B_{N+1}^{-1T}} \\
&= C_{N+1}^{-1}\,.
\end{aligned}$$

The best predicted value in the mean-square sense for X_{N+1} is Ψ. This may be seen intuitively on geometric grounds by observing that Ψ belongs to the manifold spanned by $X_0, \cdots X_N$ and $X_{N+1} - \Psi = B_{N+1}^{-1}\, \Xi_{N+1}$ is perpendicular to the manifold, so that Ψ is the orthogonal projection of X_{N+1} on the manifold. Thus, Ψ is the closest vector in the manifold to X_{N+1}.

This line of reasoning has its counterpart in the following development. Let Θ be any vector in the manifold. Then

$$\begin{aligned}
E\left\{(X_{N+1} - \Theta)\, \overline{(X_{N+1} - \Theta)^T}\right\} &= E\left\{\left[(X_{N+1} - \Psi) + (\Psi - \Theta)\right] \overline{\left[(X_{N+1} - \Psi) + (\Psi - \Theta)\right]}\right\} \\
&= E\left\{(X_{N+1} - \Psi)\, \overline{(X_{N+1} - \Psi)^T}\right\} + E\left\{(X_{N+1} - \Psi)\, \overline{(\Psi - \Theta)^T}\right\} + E\left\{(\Psi - \Theta)\, \overline{(X_{N+1} - \Psi)^T}\right\} + E\left\{(\Psi - \Theta)\, \overline{(\Psi - \Theta)^T}\right\} \\
&= E\left\{(X_{N+1} - \Psi)\, \overline{(X_{N+1} - \Psi)^T}\right\} + E\left\{(\Psi - \Theta)\, \overline{(\Psi - \Theta)^T}\right\}\,.
\end{aligned}$$

If these various nonzero expectation matrices are nonsingular then they are Hermitian symmetric and positive definite, so that for an arbitrary vector Z

$$Z^T E\left\{(X_{N+1} - \Theta)\,\overline{(X_{N+1} - \Theta)^T}\right\}Z = \overline{Z^T} E\left\{(X_{N+1} - \Psi)\,\overline{(X_{N+1} - \Psi)^T}\right\}Z + \overline{Z^T} E\{(\Psi - \Theta)\,\overline{(\Psi - \Theta)^T}\}Z$$

(where each term is real and positive), and this has its minimum at $\Theta = \Psi$.

Consequently, the multivariate linear least-square prediction of the X_{N+1} problem given $X_0, \cdots X_N$ becomes the problem of solving matrix Equation 10 for C_r where the R_{s-r} are regarded as known. The predicted value is

$$\Psi = -\sum_{r=0}^{N} B_{N+1}^{-1} B_r X_r$$

and the mean square error is $C_{N+1}^{-1} = B_{N+1}^{-1} \overline{B_{N+1}^{-1T}}$. It should be remembered that the X_m process is assumed stationary in the wide sense.

If the prediction were to be made on the basis of the entire (infinite) past, that is, on the basis of X_m where $-\infty < m < N$, then Equation 10 becomes in effect an infinite number of equations with an infinite number of unknowns and as such is unruly. In the one-dimensional case it has been found advantageous to employ the Fourier representations of Equations 2 and 3, in which case the mean square error turns out to be

$$e\left[\frac{1}{2}\int_{-1/2}^{1/2} \log F'(\lambda)\, d\lambda\right] \tag{11}$$

where $F'(\lambda)$ is the (one-dimensional) spectral density function of the process (Reference 1, p. 577). Rosenblatt (Reference 2), among others, extended the theory of prediction on an infinite past to the multivariate case but he discovered that the prediction error does not turn out to have the simple form of Equation 11 (where $F'(\lambda)$ is the matrix multivariate spectral density function) except in rather special cases. The difficulty arises from the noncommutativity of matrices in general.

The above results for prediction based on a finite past can also be obtained through use of the Fourier representation of Reference 2. Suppose that the spectral distribution function $F(\lambda)$ (an $M \times M$ matrix) is absolutely continuous so that there is a spectral density matrix $F'(\lambda)$ which we assume to be continuous and nonsingular for all values of λ. Let the spectral density matrix be taken as

$$F'(\lambda) = \left(\sum_{k=0}^{N+1} B_k\, e^{2\pi i k\lambda}\right)^{-1} \left(\sum_{k=0}^{N+1} \overline{B_k^T}\, e^{-2\pi i k\lambda}\right)^{-1}, \tag{12}$$

where

$$\left(\sum_{k=0}^{N+1} \overline{B_k^T}\, Z^{N+1-k}\right)^{-1}$$

is nonsingular for $|Z| \leqq 1$. Thus the "Hermitian square" of the matrix $H(\lambda)$ is $F'(\lambda)$:

$$H\overline{H^T} = F'(\lambda)$$

where

$$H(\lambda) = \left(\sum_{k=0}^{N+1} B_k\, e^{2\pi ik\lambda}\right)^{-1} . \tag{13}$$

Introduce this "square root" matrix into the Fourier representation of the X_m process in Equation 3 in the fashion

$$X_m = \int_{-1/2}^{1/2} e^{2\pi im\lambda}\, H(\lambda)\, d\Phi(\lambda) \tag{14}$$

where $d\Phi$ is a column matrix satisfying

$$E\left\{d\Phi(\lambda)\, \overline{d\Phi(\mu)^T}\right\} = \delta(\lambda-\mu)\, I\, d\lambda\, d\mu \ ;$$

then

$$E\left\{X_{m+n}\, \overline{X_m^T}\right\} = \int_{-1/2}^{1/2} \int_{-1/2}^{1/2} e^{2\pi i(m+n)\lambda - 2\pi im\mu}\, H(\lambda)\, \delta(\lambda-\mu)\, I\, d\lambda\, d\mu\, \overline{H(\mu)^T}$$

$$= \int_{-1/2}^{1/2} e^{2\pi in\lambda}\, H(\lambda)\overline{H(\lambda)^T} d\lambda = \int_{-1/2}^{1/2} e^{2\pi in\lambda}\, F'(\lambda)\, d\lambda \ ,$$

which agrees with Equation 2. The justification for these manipulations (including the existence of $d\Phi(\lambda)$) is discussed by Rosenblatt (Reference 2). Now we make use of the special form we have assumed for $H(\lambda)$. Define Ξ_n by

$$\Xi_{n+N+1} \equiv \sum_{k=0}^{N+1} B_k\, X_{n+k} \ ;$$

then

$$\Xi_{n+N+1} = \sum_{k=0}^{N+1} B_k \int_{-1/2}^{1/2} e^{2\pi i(n+k)\lambda} H(\lambda)\, d\Phi(\lambda)$$

$$= \int_{-1/2}^{1/2} e^{2\pi i n\lambda} \sum_{k=0}^{N+1} B_k\, e^{2\pi i k\lambda} \left(\sum_{k=0}^{N+1} B_k\, e^{2\pi i k\lambda} \right)^{-1} d\Phi(\lambda)$$

$$= \int_{-1/2}^{1/2} e^{2\pi i n\lambda}\, d\Phi(\lambda) \;,$$

which shows that the $d\Phi$ process represents Ξ_{n+N+1} just as $H\, d\Phi$ does X_n (see Equations 3 and 14). The above apparatus was set up precisely so that Ξ_n would be orthogonal to X_m for $n > m$ and so that $\{\Xi_n\}$ would be an orthonormal sequence. This gives the same orthogonal projection and linear manifold situation that was arrived at in the earlier treatment except for the greater generality now in that we are dealing with the sequence $X_n, X_{n+1}, \cdots X_{n+N+1}$ for any n instead of the one sequence $X_0, X_1, \cdots X_{N+1}$. To show the orthogonality of Ξ_n with X_m,

$$E\{\Xi_n \overline{X_m^T}\} = E\left\{ \int_{-1/2}^{1/2} e^{2\pi i(n-N-1)\lambda}\, d\Phi(\lambda) \int_{-1/2}^{1/2} e^{-2\pi i m\mu}\, \overline{\left(H(\mu)\, d\Phi(\mu)\right)^T} \right\}$$

$$= E\left\{ \int_{-1/2}^{1/2} \int_{-1/2}^{1/2} e^{2\pi i(n-N-1)\lambda - 2\pi i m\mu}\, d\Phi(\lambda)\, \overline{d\Phi(\mu)^T}\, \overline{H(\mu)^T} \right\}$$

$$= \int_{-1/2}^{1/2} \int_{-1/2}^{1/2} e^{2\pi i(n-N-1)\lambda - 2\pi i m\mu}\, \delta(\lambda - \mu)\, I\, d\lambda\, d\mu\, \overline{H(\mu)^T}$$

$$= \int_{-1/2}^{1/2} e^{2\pi i(n-m-N-1)\lambda}\, \overline{H(\lambda)^T}\, d\lambda$$

$$= \int_{-1/2}^{1/2} e^{2\pi i(n-m-N-1)\lambda} \left(\sum_{k=0}^{N+1} \overline{B_k^T}\, e^{-2\pi i k\lambda} \right)^{-1} d\lambda$$

$$= \frac{1}{2\pi i} \oint Z^{n-m-1} \left(\sum_{k=0}^{N+1} \overline{B_k^T}\, Z^{N+1-k} \right)^{-1} dZ$$

$$= \begin{cases} 0 & \text{for } n - m \geqq 1 \;, \\ \left(\overline{B_{N+1}^T}\right)^{-1} & \text{for } n = m \;. \end{cases}$$

The final integral, which is a contour integral around the unit circle, was obtained by making the transformation $e^{2\pi i\lambda} = Z$. The original defining assumption on $H(\lambda)$ in effect said that

$$\left(\sum_{k=0}^{N+1} \overline{B_k^T}\, Z^{N+1-k}\right)^{-1}$$

is analytic on and inside the unit circle. Consequently, if $n - m - 1 \geq 0$, the entire integrand is analytic on and within the contour so that the integral vanishes. On the other hand, if $n = m$ then the factor Z^{n-m-1} has a simple pole at the origin and an immediate residue calculation gives the integral's value as $(\overline{B_{N+1}^T})^{-1}$.

To show the orthonormality of Ξ_n:

$$
\begin{aligned}
E\{\Xi_n \overline{\Xi_m^T}\} &= E\left\{\int_{-1/2}^{1/2} e^{2\pi i(n-N-1)\lambda}\, d\Phi(\lambda) \int_{-1/2}^{1/2} e^{-2\pi i(m-N-1)\mu}\, \overline{d\Phi(\mu)^T}\right\} \\
&= \int_{-1/2}^{1/2}\int_{-1/2}^{1/2} e^{2\pi i(n-N-1)\lambda - 2\pi i(m-N-1)\mu}\, \delta(\lambda-\mu)\, I\, d\lambda\, d\mu \\
&= \int_{-1/2}^{1/2} e^{2\pi i(n-m)\lambda}\, I\, d\lambda \\
&= I\,\delta_{nm}\,.
\end{aligned}
$$

The matrix Equation 10 is also easily derived by the Fourier method.

$$
\begin{aligned}
\overline{\left(\sum_{k=0}^{N+1} R_{n+s-k}\, \overline{B_k^T}\right)}^{T} &= \sum_{k=0}^{N+1} B_k\, \overline{R_{n+s-k}^T} \\
&= \int_{-1/2}^{1/2} e^{-2\pi i(s+n)\lambda}\left(\sum_{k=0}^{N+1} \overline{B_k^T}\, e^{-2\pi ik\lambda}\right)^{-1} d\lambda \\
&= \sum_{k=0}^{N+1} B_k\, R_{k-s-n} \qquad \text{(by Equation 1)} \\
&= \sum_{k=0}^{N+1} B_k \int_{-1/2}^{1/2} e^{2\pi i(k-s-n)\lambda}\, F'(\lambda)\, d\lambda \\
&= \frac{1}{2\pi i}\oint Z^{N-s-n}\left(\sum_{k=0}^{N+1} \overline{B_k^T}\, Z^{N+1-k}\right)^{-1} dZ \\
&= \begin{cases} 0 & \text{for } N-s-n \geq 0\,, \\ \overline{B_{N+1}^T}^{-1} & \text{for } N-s-n = -1\,. \end{cases}
\end{aligned}
$$

Upon taking the conjugate transpose we obtain

$$\sum_{k=0}^{N+1} R_{n+s-k} \overline{B_k^T} = B_{N+1}^{-1} \delta_{s,N+1-n} \qquad \text{for } s \leqq N+1-n ,$$

which agrees with Equation 10 when $n = 0$.

PREDICTION ERROR MATRIX EQUATION

The determination of the mean square error matrix for linear least-square prediction of x_{N+1}, given the values of $x_0, \cdots x_N$ where x_m is a wide-sense stationary process, is equivalent to solving matrix Equation 10 for C_{N+1}^{-1}, which is the mean square error. This equation, while appearing to be a fairly general linear equation, is actually a bit special in that the coefficient matrices R_k are subject to the relation of Equation 1. In itself this does not appear to be of much help when it comes to a straightforward computation of C_{N+1}^{-1} from the simultaneous linear set, but further conditions on R_k may aid materially. In particular, we shall show that if R_k is Hermitian symmetric then the simultaneous set breaks up into two sets, each of approximately half the order of the original set.

Let $k = N+1-s$ in Equation 10. Then the equation is

$$\sum_{r=0}^{N+1} R_{N+1-k-r} C_r = \delta_{k0} I \qquad \text{for } 0 \leqq k \leqq N+1 . \tag{15}$$

Written out and upon application of Equation 1 this becomes

$$R_{N+1} C_0 + R_N C_1 + R_{N-1} C_2 + \cdots + R_2 C_{N-1} + R_1 C_N + R_0 C_{N+1} = I$$

$$R_N C_0 + R_{N-1} C_1 + R_{N-2} C_2 + \cdots + R_1 C_{N-1} + R_0 C_N + \overline{R_1^T} C_{N+1} = 0$$

$$R_{N-1} C_0 + R_{N-2} C_1 + R_{N-3} C_2 + \cdots + R_0 C_{N-1} + \overline{R_1^T} C_N + \overline{R_2^T} C_{N+1} = 0$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$R_2 C_0 + R_1 C_1 + R_0 C_2 + \cdots + \overline{R_{N-3}^T} C_{N-1} + \overline{R_{N-2}^T} C_N + \overline{R_{N-1}^T} C_{N+1} = 0$$

$$R_1 C_0 + R_0 C_1 + \overline{R_1^T} C_2 + \cdots + \overline{R_{N-2}^T} C_{N-1} + \overline{R_{N-1}^T} C_N + \overline{R_N^T} C_{N+1} = 0$$

$$R_0 C_0 + \overline{R_1^T} C_1 + \overline{R_2^T} C_2 + \cdots + \overline{R_{N-1}^T} C_{N-1} + \overline{R_N^T} C_N + \overline{R_{N+1}^T} C_{N+1} = 0 .$$

If we let R represent the matrix of coefficient matrices, C the matrix of the matrix unknowns, and J the column matrix of matrices on the right, then this system can be written as

$$RC = J .$$

Suppose the matrices R_m are all Hermitian symmetric. Then the matrix R takes on a good deal of symmetry, both centrosymmetry (symmetry about the center point, i.e., the point of intersection of the two main diagonals) and persymmetry (all elements of any one line perpendicular to the major diagonal alike). We shall treat only the case where N is even; the case for odd values of N can be handled similarly. Introduce the matrix D of the identity and zero submatrices:

$$D = \begin{bmatrix} I & 0 & 0 & \cdots & 0 & 0 & I \\ 0 & I & 0 & \cdots & 0 & I & 0 \\ 0 & 0 & I & \cdots & I & 0 & 0 \\ \vdots & \vdots & \vdots & \begin{smallmatrix} I & I \\ 0 & I \end{smallmatrix} & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & I & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & I & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & I \end{bmatrix},$$

and

$$D^{-1} = \begin{bmatrix} I & 0 & 0 & \cdots & 0 & 0 & -I \\ 0 & I & 0 & \cdots & 0 & -I & 0 \\ 0 & 0 & I & \cdots & -I & 0 & 0 \\ \vdots & \vdots & \vdots & \begin{smallmatrix} I & -I \\ 0 & I \end{smallmatrix} & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & I & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & I & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & I \end{bmatrix}.$$

The inverse matrix D^{-1} is to be understood in the sense that DD^{-1} is to be the matrix with the identity matrices I down the main diagonal and zero matrices elsewhere.

Now

$$D^T R D^{-1} = \begin{bmatrix} R_{N+1} & R_N & R_{N-1} & \cdots & R_{\frac{1}{2}N+1} & R_{\frac{1}{2}N} - R_{\frac{1}{2}N+1} & R_{\frac{1}{2}N-1} - R_{\frac{1}{2}N+2} & \cdots & R_2 - R_{N-1} & R_1 - R_N & R_0 - R_{N+1} \\ R_N & R_{N-1} & R_{N-2} & \cdots & R_{\frac{1}{2}N} & R_{\frac{1}{2}N-1} - R_{\frac{1}{2}N} & R_{\frac{1}{2}N-2} - R_{\frac{1}{2}N+1} & \cdots & R_1 - R_{N-2} & R_0 - R_{N-1} & R_1 - R_N \\ R_{N-1} & R_{N-2} & R_{N-3} & \cdots & R_{\frac{1}{2}N-1} & R_{\frac{1}{2}N-2} - R_{\frac{1}{2}N-1} & R_{\frac{1}{2}N-3} - R_{\frac{1}{2}N} & \cdots & R_0 - R_{N-3} & R_1 - R_{N-2} & R_2 - R_{N-1} \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ R_{\frac{1}{2}N+1} & R_{\frac{1}{2}N} & R_{\frac{1}{2}N-1} & \cdots & R_1 & R_0 - R_1 & R_1 - R_2 & \cdots & R_{\frac{1}{2}N-2} - R_{\frac{1}{2}N-1} & R_{\frac{1}{2}N-1} - R_{\frac{1}{2}N} & R_{\frac{1}{2}N} - R_{\frac{1}{2}N+1} \\ R_{\frac{1}{2}N} + R_{\frac{1}{2}N+1} & R_{\frac{1}{2}N-1} + R_{\frac{1}{2}N} & R_{\frac{1}{2}N-2} + R_{\frac{1}{2}N-1} & \cdots & R_0 + R_1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ R_{\frac{1}{2}N-1} + R_{\frac{1}{2}N+2} & R_{\frac{1}{2}N-2} + R_{\frac{1}{2}N+1} & R_{\frac{1}{2}N-3} + R_{\frac{1}{2}N} & \cdots & R_1 + R_2 & 0 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ R_2 + R_{N-1} & R_1 + R_{N-2} & R_0 + R_{N-3} & \cdots & R_{\frac{1}{2}N-2} + R_{\frac{1}{2}N-1} & 0 & 0 & \cdots & 0 & 0 & 0 \\ R_1 + R_N & R_0 + R_{N-1} & R_1 + R_{N-2} & \cdots & R_{\frac{1}{2}N-1} + R_{\frac{1}{2}N} & 0 & 0 & \cdots & 0 & 0 & 0 \\ R_0 + R_{N+1} & R_1 + R_N & R_2 + R_{N-1} & \cdots & R_{\frac{1}{2}N} + R_{\frac{1}{2}N+1} & 0 & 0 & \cdots & 0 & 0 & 0 \end{bmatrix},$$

$$
DC = \begin{bmatrix} C_0 + C_{N+1} \\ C_1 + C_N \\ C_2 + C_{N-1} \\ \vdots \\ C_{\frac{1}{2}N} + C_{\frac{1}{2}N+1} \\ C_{\frac{1}{2}N+1} \\ \vdots \\ C_{N-1} \\ C_N \\ C_{N+1} \end{bmatrix} ,
$$

and

$$
D^T J = \begin{bmatrix} I \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ I \end{bmatrix} .
$$

The equation $RC = J$ is equivalent to $D^T RD^{-1} DC = D^T J$ and in view of the three matrices just calculated this is equivalent to the two equations

$$
\begin{bmatrix}
R_{\frac{1}{2}N} + R_{\frac{1}{2}N+1} & R_{\frac{1}{2}N-1} + R_{\frac{1}{2}N} & R_{\frac{1}{2}N-2} + R_{\frac{1}{2}N-1} & \cdots & R_0 + R_1 \\
R_{\frac{1}{2}N-1} + R_{\frac{1}{2}N+2} & R_{\frac{1}{2}N-2} + R_{\frac{1}{2}N+1} & R_{\frac{1}{2}N+3} + R_{\frac{1}{2}N} & \cdots & R_1 + R_2 \\
\vdots & \vdots & \vdots & & \vdots \\
R_2 + R_{N-1} & R_1 + R_{N-2} & R_0 + R_{N-3} & \cdots & R_{\frac{1}{2}N-2} + R_{\frac{1}{2}N-1} \\
R_1 + R_N & R_0 + R_{N-1} & R_1 + R_{N-2} & \cdots & R_{\frac{1}{2}N-1} + R_{\frac{1}{2}N} \\
R_0 + R_{N+1} & R_1 + R_N & R_2 + R_{N-1} & \cdots & R_{\frac{1}{2}N} + R_{\frac{1}{2}N+1}
\end{bmatrix}
\begin{bmatrix}
C_0 + C_{N+1} \\ C_1 + C_N \\ \vdots \\ C_{\frac{1}{2}N-2} + C_{\frac{1}{2}N+3} \\ C_{\frac{1}{2}N-1} + C_{\frac{1}{2}N+2} \\ C_{\frac{1}{2}N} + C_{\frac{1}{2}N+1}
\end{bmatrix}
=
\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ I \end{bmatrix} , \qquad (16a)
$$

and

$$\begin{bmatrix} R_{N+1} & R_N & R_{N-1} & \cdots & R_{\frac{1}{2}N+1} & R_{\frac{1}{2}N} - R_{\frac{1}{2}N+1} & R_{\frac{1}{2}N-1} - R_{\frac{1}{2}N+2} & \cdots & R_2 - R_{N-1} & R_1 - R_N & R_0 - R_{N+1} \\ R_N & R_{N-1} & R_{N-2} & \cdots & R_{\frac{1}{2}N} & R_{\frac{1}{2}N-1} - R_{\frac{1}{2}N} & R_{\frac{1}{2}N-2} - R_{\frac{1}{2}N+1} & \cdots & R_1 - R_{N-2} & R_0 - R_{N-1} & R_1 - R_N \\ R_{N-1} & R_{N-2} & R_{N-3} & \cdots & R_{\frac{1}{2}N-1} & R_{\frac{1}{2}N-2} - R_{\frac{1}{2}N-1} & R_{\frac{1}{2}N-3} - R_{\frac{1}{2}N} & \cdots & R_0 - R_{N-3} & R_1 - R_{N-2} & R_2 - R_{N-1} \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ R_{\frac{1}{2}N+1} & R_{\frac{1}{2}N} & R_{\frac{1}{2}N-1} & \cdots & R_1 & R_0 - R_1 & R_1 - R_2 & \cdots & R_{\frac{1}{2}N-2} - R_{\frac{1}{2}N-1} & R_{\frac{1}{2}N-1} - R_{\frac{1}{2}N} & R_{\frac{1}{2}N} - R_{\frac{1}{2}N+1} \end{bmatrix} \cdot \begin{bmatrix} C_0 + C_{N+1} \\ C_1 + C_N \\ C_2 + C_{N-1} \\ \vdots \\ C_{\frac{1}{2}N} + C_{\frac{1}{2}N+1} \\ C_{\frac{1}{2}N+1} \\ C_{\frac{1}{2}N+2} \\ \vdots \\ C_{N+1} \end{bmatrix} = \begin{bmatrix} I \\ 0 \\ 0 \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ 0 \end{bmatrix} \tag{16b}$$

The last equation can be written as

$$\begin{bmatrix} R_{\frac{1}{2}N} - R_{\frac{1}{2}N+1} & R_{\frac{1}{2}N-1} - R_{\frac{1}{2}N+2} & \cdots & R_2 - R_{N-1} & R_1 - R_N & R_0 - R_{N+1} \\ R_{\frac{1}{2}N-1} - R_{\frac{1}{2}N} & R_{\frac{1}{2}N-2} - R_{\frac{1}{2}N+1} & \cdots & R_1 - R_{N-2} & R_0 - R_{N-1} & R_1 - R_N \\ R_{\frac{1}{2}N-2} - R_{\frac{1}{2}N-1} & R_{\frac{1}{2}N-3} - R_{\frac{1}{2}N} & \cdots & R_0 - R_{N-3} & R_1 - R_{N-2} & R_2 - R_{N-1} \\ \vdots & \vdots & & \vdots & \vdots & \vdots \\ R_0 - R_1 & R_1 - R_2 & & R_{\frac{1}{2}N-2} - R_{\frac{1}{2}N-1} & R_{\frac{1}{2}N-1} - R_{\frac{1}{2}N} & R_{\frac{1}{2}N} - R_{\frac{1}{2}N+1} \end{bmatrix} \begin{bmatrix} C_{\frac{1}{2}N+1} \\ C_{\frac{1}{2}N+2} \\ C_{\frac{1}{2}N+3} \\ \vdots \\ C_{N+1} \end{bmatrix} = \begin{bmatrix} I \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} - \begin{bmatrix} R_{N+1} & R_N & \cdots & R_{\frac{1}{2}N+1} \\ R_N & R_{N-1} & \cdots & R_{\frac{1}{2}N} \\ R_{N-1} & R_{N-2} & \cdots & R_{\frac{1}{2}N-1} \\ \vdots & \vdots & & \vdots \\ R_{\frac{1}{2}N+1} & R_{\frac{1}{2}N} & \cdots & R_1 \end{bmatrix} \begin{bmatrix} C_0 + C_{N+1} \\ C_1 + C_N \\ C_2 + C_{N-1} \\ \vdots \\ C_{\frac{1}{2}N} + C_{\frac{1}{2}N+1} \end{bmatrix}$$

where the $C_0 + C_{N+1}$ column matrix would have been found from Equation 16a. Thus, if R_n is Hermitian symmetric the original system of equations can be solved by solving successively two systems each of which is about half the order of the original.

Naturally, if further conditions are imposed on R_n further results may be obtained. The next section is devoted to quite special choices for R_n with a study of what can be said about C_{N+1}^{-1} in these cases. Before proceeding to these, though, we give one other example.

Suppose $R_n = \alpha_n K$ where K is a nonsingular Hermitian symmetric constant matrix and α_n is a scalar function of n satisfying $\alpha_{-n} = \alpha_n$ in accordance with Equation 1. The fundamental Equation 15 then becomes

$$\sum_{r=0}^{N+1} \alpha_{N+1-k-r} K C_r = \delta_{k0} I \qquad \text{for } 0 \leqq k \leqq N+1 ,$$

which after the substitution $X_r = KC_r$ is

$$\sum_{r=0}^{N+1} \alpha_{N+1-k-r} X_r = \delta_{k0} I \qquad \text{for } 0 \leqq k \leqq N+1 .$$

Write X_r in the form $X_r = \lambda_r I + Y_r$ where λ_r is a scalar function of r so chosen that

$$\sum_{r=0}^{N+1} \alpha_{N+1-k-r} \lambda_r = \delta_{k0} \qquad \text{for } 0 \leqq k \leqq N+1 . \tag{17}$$

The matrix of coefficients of λ_r is assumed nonsingular. Then we find

$$\sum_{r=0}^{N+1} \alpha_{N+1-k-r} \left(\lambda_r \, I + Y_r\right) = I \sum_{r=0}^{N+1} \alpha_{N+1-k-r} \, \lambda_r + \sum_{r=0}^{N+1} \alpha_{N+1-k-r} \, Y_r$$

$$= I \, \delta_{k0} + \sum_{r=0}^{N+1} \alpha_{N+1-k-r} \, Y_r \, ,$$

which means that

$$\sum_{r=0}^{N+1} \alpha_{N+1-k-r} \, Y_r = 0$$

and therefore $Y_r = 0$. From this it follows that $C_r = \lambda_r K^{-1}$. Let

$$\Delta_{N+1} \equiv \begin{vmatrix} \alpha_{N+1} & \alpha_N & \alpha_{N-1} & \cdots & \alpha_1 & \alpha_0 \\ \alpha_N & \alpha_{N-1} & \alpha_{N-2} & \cdots & \alpha_0 & \alpha_1 \\ \alpha_{N-1} & \alpha_{N-2} & \alpha_{N-3} & \cdots & \alpha_1 & \alpha_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \alpha_0 & \alpha_1 & \alpha_2 & \cdots & \alpha_N & \alpha_{N+1} \end{vmatrix} ,$$

a centrosymmetric persymmetric determinant. Then the system of Equations 17 can be solved for λ_r, and λ_{N+1} has the form

$$\lambda_{N+1} = (-1)^{N+1} \frac{\Delta_N}{\Delta_{N+1}} \, .$$

Consequently, the mean square error is simply

$$C_{N+1}^{-1} = (-1)^{N+1} \frac{\Delta_{N+1}}{\Delta_N} \, K \, .$$

EXAMPLES

Phase Space Process

Regard x_t as a one-dimensional continuous parameter process and denote its derivative by $\dot{x}_t$. The x_t process has the Fourier representation (Reference 1)

$$x_t = \int_{-\infty}^{\infty} e^{2\pi i \lambda t} \, dZ(\lambda)$$

so that the formal derivative $\dot{x}_t$ is given by

$$\dot{x}_t = \int_{-\infty}^{\infty} 2\pi i\lambda \, e^{2\pi i\lambda t} \, dZ(\lambda) \quad .$$

If we were to sample these at equal intervals we would expect these to go over to

$$x_n = \int_{-1/2}^{1/2} e^{2\pi i\lambda n} \, dy(\lambda)$$

and

$$\dot{x}_n = \int_{-1/2}^{1/2} 2\pi i\lambda \, e^{2\pi i\lambda n} \, dy(\lambda) \quad .$$

Thus we are led to define the two-dimensional "phase space process" X_n as

$$X_n = \begin{bmatrix} x_n \\ \dot{x}_n \end{bmatrix} = \int_{-1/2}^{1/2} e^{2\pi i\lambda n} \begin{bmatrix} dy(\lambda) \\ 2\pi i\lambda \, dy(\lambda) \end{bmatrix} \equiv \int_{-1/2}^{1/2} e^{2\pi i\lambda n} \, dY(\lambda) \quad , \tag{18}$$

where

$$dY = \begin{bmatrix} 1 \\ 2\pi i\lambda \end{bmatrix} dy(\lambda) \equiv \Lambda \, dy(\lambda) \quad .$$

To obtain the covariance matrix of X_n we determine the expected value of $dY \, \overline{dY^T}$:

$$E\{dY \, \overline{dY^T}\} = E\{\Lambda \, dy \, \overline{\Lambda^T} \, \overline{dy}\} = \Lambda\overline{\Lambda^T} \, E\{dy \, \overline{dy}\} \equiv P(\lambda) \, f'(\lambda) \, \delta(\lambda - \mu) \, d\lambda \, d\mu \quad ,$$

where

$$P(\lambda) \equiv \Lambda\overline{\Lambda^T} = \begin{bmatrix} 1 & -2\pi i\lambda \\ 2\pi i\lambda & 4\pi^2\lambda^2 \end{bmatrix} .$$

Here $f'(\lambda)$ is the spectral density function for the one-dimensional x_n process. Consequently,

$$R_m = E\{X_{m+n} \, \overline{X_n^T}\} = \int_{-1/2}^{1/2} e^{2\pi i m\lambda} \, P(\lambda) \, df(\lambda) \equiv \int_{-1/2}^{1/2} e^{2\pi i m\lambda} \, dF(\lambda) \quad ; \tag{19}$$

that is, the spectral density function $F'(\lambda)$ of the X_m process is equal to that of the x_m process multiplied by the matrix $P(\lambda)$. It is of interest to determine the counterparts of the above formulae in the real case. As seen in the work leading to Equation 5 the one-dimensional process will be real if $dy = dy_1 + i dy_2$ with $dy_1(\lambda)$ being an even function of λ and $dy_2(\lambda)$, an odd. Then, upon letting $du = 2dy_1$ and $dv = -2dy_2$, we obtain, as in Equation 5,

$$x_n = \int_0^{1/2} (\cos 2\pi n\lambda)\, du(\lambda) + (\sin 2\pi n\lambda)\, dv(\lambda)$$

and

$$\dot{x}_n = \int_0^{1/2} -2\pi\lambda (\sin 2\pi n\lambda)\, du + 2\pi\lambda (\cos 2\pi n\lambda)\, dv \,.$$

This is representable in the two-dimensional form as

$$X_n = \int_0^{1/2} (\cos 2\pi n\lambda)\, dU(\lambda) + (\sin 2\pi n\lambda)\, dV(\lambda) \tag{20}$$

where

$$dU = \begin{bmatrix} du \\ 2\pi\lambda\, dv \end{bmatrix}$$

and

$$dV = \begin{bmatrix} dv \\ -2\pi\lambda\, du \end{bmatrix} . \tag{21}$$

The du and dv satisfy Equation 7,

$$E\{du(\lambda)\, du(\mu)\} = E\{dv(\lambda)\, dv(\mu)\} = g'(\lambda)\, \delta(\lambda-\mu)\, d\lambda\, d\mu \,,$$

$$E\{du(\lambda)\, dv(\mu)\} = E\{dv(\lambda)\, du(\mu)\} = 0 \,,$$

but dU and dV as given by Equation 21 do not satisfy Equation 7. In fact,

$$E\{dU(\lambda)\, dU^T_{(\mu)}\} = E\{dV\, dV^T_{(\mu)}\} = A\, g'(\lambda)\, \delta(\lambda-\mu)\, d\lambda\, d\mu \,,$$

$$E\{dU(\lambda)\, dV^T_{(\mu)}\} = -E\{dV(\lambda)\, dU^T_{(\mu)}\} = B\, g'(\lambda)\, \delta(\lambda-\mu)\, d\lambda\, d\mu \,,$$

where

$$A \quad \begin{bmatrix} 1 & 0 \\ 0 & 4\pi^2\lambda^2 \end{bmatrix}$$

and

$$B \quad \begin{bmatrix} 0 & -2\pi\lambda \\ 2\pi\lambda & 0 \end{bmatrix}$$

Consequently,

$$R_m \quad E\left\{X_{m+n} X_n^T\right\} = \int_0^{1/2}\int_0^{1/2} \left(\cos 2\pi (m+n)\lambda\right)(\cos 2\pi n\mu)\, E\left\{dU\, dU^T\right\} + \left(\cos 2\pi(m+n)\lambda\right)(\sin 2\pi n\mu)\, E\left\{dU\, dV^T\right\}$$

$$+ \left(\sin 2\pi(m+n)\lambda\right)(\cos 2\pi n\mu)\, E\left\{dV\, dU^T\right\} + \left(\sin 2\pi(m+n)\lambda\right)(\sin 2\pi n\mu)\, E\left\{dV\, dV^T\right\}$$

$$\int_0^{1/2} (\cos 2\pi m\lambda)\, A\, g'(\lambda)\, d\lambda - (\sin 2\pi m\lambda)\, B\, g'(\lambda)\, d\lambda \quad ,$$

or, when written out explicitly,

$$R_m \quad \begin{bmatrix} \displaystyle\int_0^{1/2} \cos 2\pi m\lambda\, g'(\lambda)\, d\lambda & \displaystyle 2\pi \int_0^{1/2} \lambda \sin 2\pi m\lambda\, g'(\lambda)\, d\lambda \\ \displaystyle -2\pi \int_0^{1/2} \lambda \sin 2\pi m\lambda\, g'(\lambda)\, d\lambda & \displaystyle 4\pi^2 \int_0^{1/2} \lambda^2 \cos 2\pi m\lambda\, g'(\lambda)\, d\lambda \end{bmatrix} . \tag{22}$$

Observe from Equation 22 that one general characteristic of a phase space process is that the off-diagonal elements of the covariance matrix are negatives of each other. Also, R_m^{21} is the derivative of R_m^{11} with respect to m and R_m^{22} is the negative of the second derivative of R_m^{11}.

If $g(\lambda)$ is a step function, so that $g'(\lambda)$ has the form

$$g'(\lambda) = \sum_{j=1}^{k} \sigma_j^2\, \delta\left(\lambda - \lambda_j\right) \quad ,$$

then Equation 22 becomes a sum of matrices

$$R_m = \sum_{j=1}^{k} \sigma_j^2 \begin{bmatrix} \cos 2\pi m\lambda_j & 2\pi\lambda_j \sin 2\pi m\lambda_j \\ -2\pi\lambda_j \sin 2\pi m\lambda_j & +\pi^2 \lambda_j^2 \cos 2\pi m\lambda_j \end{bmatrix} . \tag{23}$$

It may be noted that the matrix $P(\lambda)$ in Equation 19 is singular and so, therefore, is $F'(\lambda)$. Thus, we are not in a position to apply the outlined Fourier approach starting with Equation 12 for prediction on a finite past for a phase space process.

Degenerate Quasi Phase Space Process

Let the covariance matrix have a value approximately that of Equation 23 with $k = 1$. That is,

$$R_m = \begin{bmatrix} \sigma^2 \cos(2\pi m\lambda) + \epsilon\rho_m^{11} & 2\pi\lambda\sigma^2 \sin 2\pi m\lambda + \epsilon\rho_m^{12} \\ -2\pi\lambda\sigma^2 \sin 2\pi m\lambda + \epsilon\rho_m^{21} & 4\pi^2 \lambda^2 \sigma^2 \cos 2\pi m\lambda + \epsilon\rho_m^{22} \end{bmatrix}$$

$$\equiv P_m + \epsilon\rho_m \qquad \text{with} \quad \rho_{-m} = \rho_m^T , \qquad \epsilon \ll 1 .$$

We shall show that in this case the mean square prediction error is small; it is, in fact, of the order of ϵ.

The demonstration makes fundamental use of the relation

$$P_n P_0^{-1} P_m^T = P_{n-m} \tag{24}$$

which follows from a simple matrix calculation.

We use the equation connecting the prediction error with the covariance matrix in the form

$$\sum_{r=0}^{N+1} R_{N+1-p-r} C_r = I \delta_{p0} \qquad \text{for} \quad 0 \leqq p \leqq N+1 . \tag{25}$$

Take the convolution of this with the resolvent kernel

$$c_{np} = \begin{cases} \delta_{np} I - P_{N+1-n} P_0^{-1} \delta_{N+1,p} & \text{for } 0 \leqq n \leqq N , \quad 0 \leqq p \leqq N+1 , \\ P_0^{-1} \delta_{N+1,p} & \text{for } n = N+1, \quad 0 \leqq p \leqq N+1 , \end{cases}$$

and get

$$\sum_{p=0}^{N+1} c_{np} \sum_{r=0}^{N+1} \left(P_{N+1-p-r} + \epsilon \rho_{N+1-p-r}\right) C_r = \sum_{p=0}^{N+1} c_{np}\ I\ \delta_{p0}\ ,$$

$$\sum_{r=0}^{N+1} \left[\sum_{p=0}^{N+1} c_{np} \left(P_{N+1-p-r} + \epsilon \rho_{N+1-p-r}\right)\right] C_r = \delta_{n0}\ I\ .$$

Now, for $0 \leqq n \leqq N$,

$$\begin{aligned}\sum_{p=0}^{N+1} c_{np}\ P_{N+1-p-r} &= \sum_{p=0}^{N+1} \left(\delta_{np}\ I - P_{N+1-n} P_0^{-1}\ \delta_{N+1,p}\right) P_{N+1-p-r} \\ &= P_{N+1-n-r} - P_{N+1-n}\ P_0^{-1}\ P_{-r} \\ &= P_{N+1-n-r} - P_{N+1-n}\ P_0^{-1}\ P_r^T = 0\end{aligned}$$

from Equation 24. Again, for $0 \leqq n \leqq N$,

$$\epsilon \sum_{p=0}^{N+1} c_{np}\ \rho_{N+1-p-r} = \epsilon\left(\rho_{N+1-n-r} - P_{N+1-n}\ P_0^{-1}\ \rho_r^T\right)\ ,$$

whereas for $n = N + 1$ we obtain

$$\sum_{p=0}^{N+1} c_{N+1,p}\, P_{N+1-p-r} = \sum_{p=0}^{N+1} P_0^{-1}\ \delta_{N+1,p}\, P_{N+1-p-r} = P_0^{-1}\ P_r^T\ ,$$

and

$$\epsilon \sum_{p=0}^{N+1} c_{N+1,p}\ \rho_{N+1-p-r} = \epsilon P_0^{-1}\ \rho_r^T\ .$$

Consequently, the result of taking the convolution of Equation 15 with c_{np} is

$$\left.\begin{aligned} 0 + \epsilon \sum_{r=0}^{N+1} \left(\rho_{N+1-n-r} - P_{N+1-n}\ P_0^{-1}\ \rho_r^T\right) C_r &= \delta_{n0}\ I && \text{for } 0 \leqq n \leqq N\ , \\ \sum_{r=0}^{N+1} P_0^{-1}\ P_r^T\ C_r + \epsilon \sum_{r=0}^{N+1} P_0^{-1}\ \rho_r^T\ C_r &= 0 && \text{for } n = N+1\ . \end{aligned}\right\} \quad (26)$$

The first of these equations is of the form $\epsilon PC = J$ where P and J are independent of ϵ and the second equation upon multiplication by ϵ is of the form $\epsilon(\hat{P} + \epsilon\hat{\hat{P}})\,C = 0$ with $\hat{P}$, $\hat{\hat{P}}$ independent of ϵ. If $\epsilon \to 0$, ϵC must remain bounded away from zero since the right-hand side of Equation 26 stays fixed and nonzero. This implies that $C_{N+1}^{-1} = O(\epsilon)$, if C_{N+1}^{-1} exists (which will be the case for general values of ρ_m).

The fact that the convolution of $P_{N+1-p-r}$ with c_{np} vanishes for $0 \leqq n \leqq N$ is connected with the fact that the matrix $P(\lambda)$ in Equation 19 is singular.

Quasi Cosine Process

Take the covariance matrix to be

$$R_m = \cos 2\pi m\lambda\, Q_0 + \epsilon\rho_m \qquad \text{with } \rho_{-m} = \rho_m^T , \qquad \epsilon \ll 1 , \qquad 0 < \lambda < \pi ,$$

where Q_0 is a nonsingular symmetric constant matrix (i.e., independent of m). At first glance this covariance matrix looks simpler than that of the example for the degenerate quasi phase space process but the resolvent kernel is more involved. It is, in fact,

$$b_{np} = \begin{cases} \left\{\delta_{np} - \left[\sin 2\pi (N+1-n)\lambda\right](\csc 2\pi\lambda)\,\delta_{Np} + \left[\sin 2\pi (N-n)\lambda\right](\csc 2\pi\lambda)\,\delta_{N+1,p}\right\} I & \text{for } 0 \leqq n \leqq N-1 , \\ \left(\csc^2 2\pi\lambda\, Q_0^{-1}\right)\delta_{Np} - (\csc 2\pi\lambda)\left(\cot 2\pi\lambda\right) Q_0^{-1}\,\delta_{N+1,p} & \text{for } n = N , \\ Q_0^{-1}\,\delta_{N+1,p} & \text{for } n = N+1 . \end{cases}$$

The crucial computation is a bit of trigonometry showing that for $0 \leqq n \leqq N-1$

$$\sum_{p=0}^{N+1} b_{np} \cos 2\pi (N+1-p-r)\lambda = 0 .$$

The result of convolving b_{np} with Equation 15 is

$$\begin{cases} 0 + \epsilon \displaystyle\sum_{r=0}^{N+1}\left\{\rho_{N+1-n-r} - \left[\sin 2\pi (N+1-n)\lambda\right](\csc 2\pi\lambda)\,\rho_{1-r} + \left[\sin 2\pi (N-n)\lambda\right](\csc 2\pi\lambda)\,\rho_r^T\right\} C_r = \delta_{n0}\, I \\ \qquad\qquad \text{for } 0 \leqq n \leqq N-1 , \\ \displaystyle\sum_{r=0}^{N+1} (\csc 2\pi\lambda)(\sin 2\pi r\lambda)\, C_r + \epsilon \sum_{r=0}^{N+1} (\csc 2\pi\lambda)\left[(\csc 2\pi\lambda)\,\rho_{1-r} - (\cos 2\pi\lambda)\,\rho_r^T\right] Q_0^{-1}\, C_r = 0 \qquad \text{for } n = N , \\ \displaystyle\sum_{r=0}^{N+1} (\cos 2\pi r\lambda)\, C_r + \epsilon \sum_{r=0}^{N+1} \rho_r^T Q_0^{-1}\, C_r = 0 \qquad \text{for } n = N+1 , \end{cases}$$

and reasoning as in the previous example leads us to the same conclusion; i.e., that here also $C_{N+1}^{-1} = O(\epsilon)$.

Quasi White Noise Process

Take the covariance matrix to be

$$R_m = \delta_{m0} Q_0 + \epsilon \rho_m \qquad \text{with } \rho_{-m} = \rho_m^T , \quad \epsilon \ll 1 ,$$

where again Q_0 is a nonsingular symmetric constant matrix. This is a highly degenerate case and fairly strong results can be obtained. The equation to be solved is

$$\sum_{r=0}^{N+1} \left(\delta_{N+1-p-r,0} Q_0 + \epsilon \rho_{N+1-p-r}\right) C_r = I \delta_{p0} \qquad \text{for } 0 \leqq p \leqq N+1 . \tag{27}$$

If ϵ is set equal to zero it is easily verified that the solution is $C_r = 0$, $0 \leqq r \leqq N$, $C_{N+1} = Q_0^{-1}$. Thus, if we think of the solution C_r to Equation 27 as a function of ϵ, $C_r = C_r(\epsilon)$, then $C_r(\epsilon) = o(1)$, $0 \leqq r \leqq N$, and $C_{N+1}(\epsilon) = O(1)$ as $\epsilon \to 0$. Now take the convolution of Equation 27 with the kernel

$$e_{np}^{1} = \begin{cases} \delta_{np} I - \epsilon \rho_{N+1-n} Q_0^{-1} \delta_{N+1,p} & \text{for } 0 \leqq n \leqq N , \quad 0 \leqq p \leqq N+1 , \\ \delta_{N+1,p} I & \text{for } n = N+1 , \quad 0 \leqq p \leqq N+1 , \end{cases}$$

and obtain

$$\sum_{r=0}^{N+1} \left[\delta_{N+1-n-r,0} Q_0 + \epsilon \left(1 - \delta_{r,0}\right) \rho_{N+1-n-r} + O\left(\epsilon^2\right)\right] C_r = \delta_{n0} I \qquad \text{for } 0 \leqq n \leqq N ,$$

$$\sum_{r=0}^{N+1} \left(\delta_{n0} Q_0 + \epsilon \rho_r^T\right) C_r = 0 \qquad \text{for } n = N+1 .$$

Observe that through the terms in ϵ^0 and ϵ^1 the first of these equations is of the same form as Equation 27 except that the ranges of r and p have both been reduced by one. This is made even more apparent by rewriting the equation as

$$O\left(\epsilon^2\right) C_0 + \sum_{r=1}^{N+1} \left[\delta_{N+2-n-r,0} Q_0 + \epsilon \rho_{N+2-p-r} + O\left(\epsilon^2\right)\right] C_r = \delta_{p1} I \qquad \text{for } 1 \leqq p \leqq N+1 ,$$

wherein the transformation $n = p+1$ has been made to make the ranges of r and p the same. This suggests a reapplication of the above convolution procedure. The convolution kernel of choice now is

$$e^2_{np} = \begin{cases} \delta_{np}\, I - \epsilon\rho_{N+1-n}\, Q_0^{-1}\, \delta_{N+1,p} & \text{for } 1 \leqq n \leqq N\,, \quad 1 \leqq p \leqq N+1\,, \\ \delta_{N+1,p}\, I & \text{for } n = N+1\,, \quad 1 \leqq p \leqq N+1\,. \end{cases}$$

Then upon taking the convolution as before we get

$$O(\epsilon^2)\, C_0 + \sum_{r=1}^{N+1} \left[\delta_{N+2-n-r,0}\, Q_0 + \epsilon\left(1 - \delta_{r1}\right) \rho_{N+2-n-r} + O(\epsilon^2)\right] C_r = \delta_{n1} I \qquad \text{for } 1 \leqq n \leqq N\,,$$

$$O(\epsilon^2)\, C_0 + \sum_{r=1}^{N+1} \left[\delta_{r1}\, Q_0 + \epsilon\rho_{1-r} + O(\epsilon^2)\right] C_r = 0 \qquad \text{for } n = N+1\,.$$

By a little manipulation the first of these can be thrown into the form

$$O(\epsilon^2)\, C_0 + O(\epsilon^2)\, C_1 + \sum_{r=2}^{N+1} \left[\delta_{N+3-p-r,0}\, Q_0 + \epsilon\rho_{N+3-p-r} + O(\epsilon^2)\right] C_r = \delta_{p2}\, I \qquad \text{for } 2 \leqq p \leqq N+1\,.$$

We continue in this fashion and ultimately obtain

$$O(\epsilon^2)\, C_0 + O(\epsilon^2)\, C_1 + \cdots + O(\epsilon^2)\, C_N + \sum_{r=N+1}^{N+1} \left[\delta_{N+N+2-p-r,0}\, Q_0 + \epsilon\rho_{N+N+2-p-r} + O(\epsilon^2)\right] C_r = \delta_{p,N+1}\, I \qquad \text{for } p = N+1\,,$$

or

$$O(\epsilon^2)\, C_0 + O(\epsilon^2)\, C_1 + \cdots + O(\epsilon^2)\, C_N + \left[Q_0 + \epsilon\rho_0 + O(\epsilon^2)\right] C_{N+1} = I\,.$$

Now as seen earlier $C_r = o(1)$ for $0 \leqq r \leqq N$ as $\epsilon \to 0$. Hence we have

$$\left[Q_0 + \epsilon\rho_0 + O(\epsilon^2)\right] C_{N+1} = I + o(\epsilon^2)\,,$$

and this may be seen to be equivalent with

$$C_{N+1}^{-1} = Q_0 + \epsilon\rho_0 + O(\epsilon^2)\,. \tag{28}$$

Exponential Decay Process

Take the covariance matrix to be

$$R_m = e^{-a|m|} Q_0 \equiv V_m ,$$

where a is positive and real and Q_0 is a real nonsingular symmetric constant matrix. This covariance matrix is of the form $\alpha_m K$ which was investigated earlier in the section entitled "Prediction Error Matrix Equation." We saw in that case that solving the matrix equation for C_{N+1}^{-1} could be accomplished by evaluating two determinants defined on α_m. However, we shall not use that result now but instead will obtain C_{N+1}^{-1} by the methods employed in the other examples. The equation to be treated is

$$\sum_{r=0}^{N+1} e^{-a|N+1-p-r|} Q_0 C_r = I \delta_{p0} \qquad \text{for } 0 \leqq p \leqq N+1 ,$$

and the kernel that suffices is

$$v_{np} = \begin{cases} \delta_{np} I - \delta_{n+1,p} e^{-a} I & \text{for } 0 \leqq n \leqq N , \quad 0 \leqq p \leqq N+1 , \\ \delta_{N+1,p} I & \text{for } n = N+1, \quad 0 \leqq p \leqq N+1 . \end{cases}$$

In fact,

$$\begin{aligned} \sum_{p=0}^{N+1} v_{np} e^{-a|N+1-p-r|} &= \sum_{p=0}^{N+1} \left(\delta_{np} I - \delta_{n+1,p} e^{-a} I\right) e^{-a|N+1-p-r|} \qquad \text{for } 0 \leqq n \leqq N , \\ &= I e^{-a|N+1-n-r|} - I e^{-a-a|N-n-r|} \\ &= I e^{-a|N+1-n-r|} \left(1 - e^{a|N+1-n-r|-a|N-n-r|-a}\right) \\ &= \begin{cases} I e^{-a|N+1-n-r|} \left(1 - e^{0}\right) & \text{for } N-n-r \geqq 0 , \quad 0 \leqq n \leqq N , \\ I e^{-a|N+1-n-r|} \left(1 - e^{-2a}\right) & \text{for } N-n-r < 0 , \quad 0 \leqq n \leqq N . \end{cases} \end{aligned}$$

The convolved equation is therefore the diagonal system

$$\sum_{r=N+1-n}^{N+1} I e^{-a|N+1-n-r|} \left(1 - e^{-2a}\right) Q_0 C_r = \delta_{n0} I \qquad \text{for } 0 \leqq n \leqq N ,$$

$$\sum_{r=0}^{N+1} I e^{-ar} Q_0 C_r = 0 \qquad \text{for } n = N+1 .$$

All we need do is take $n = 0$, as this equation is merely

$$I e^{0} \left(1 - e^{-2a}\right) Q_0 C_{N+1} = I ,$$

which gives

$$C_{N+1}^{-1} = \left(1 - e^{-2a}\right) Q_0 .$$

Quadratic Exponential Decay Process

Take the covariance matrix to be

$$R_m = e^{-an^2+bn} Q_0 = U_m ,$$

where a is positive and real, b is pure imaginary, and Q_0 is a nonsingular Hermitian symmetric constant matrix. These qualifications are imposed to insure that the covariance matrix fulfills Equation 1. This example is most simply treated by applying a succession of convolutions, the first being with the kernel

$$u_{np}^{1} = \begin{cases} \delta_{np} I - \delta_{n+1,p} e^{-a(2N+1-2n)+b} I & \text{for } 0 \leqq n \leqq N , \\ \delta_{N+1,p} I & \text{for } n = N+1 . \end{cases}$$

This convolution produces the system

$$\sum_{r=0}^{N+1} I e^{-a(N+1-n-r)^2+b(N+1-n-r)} \left(1 - e^{-2ar}\right) Q_0 C_r = \delta_{n0} I \qquad \text{for } 0 \leqq n \leqq N ,$$

$$\sum_{r=0}^{N+1} I e^{-ar^2-br} Q_0 C_r = 0 \qquad \text{for } n = N+1 .$$

But the factor $1 - e^{-2ar}$ vanishes for $r = 0$; so we now have a reduced system to deal with:

$$\sum_{r=1}^{N+1} I e^{-a(N+2-p-r)^2+b(N+2-p-r)} Q_0 \left(1 - e^{-2ar}\right) C_r = \delta_{p1} I \qquad \text{for } 1 \leqq p \leqq N+1 .$$

This is of the same general type as the original equation and so we proceed as before, the kernel now being

$$u_{np}^2 = \begin{cases} \delta_{np} I - \delta_{n+1,p}\, e^{-a(2N+1-2n)+b} I & \text{for } 1 \leqq n \leqq N, \quad 1 \leqq p \leqq N+1, \\ \delta_{N+1,p} I & \text{for } n = N+1, \quad 1 \leqq p \leqq N+1, \end{cases}$$

with the resulting convolved equation

$$\sum_{r=1}^{N+1} I\, e^{-a(N+2-n-r)^2+b(N+2-n-r)} Q_0 \left(1 - e^{-2ar}\right)\left(1 - e^{-2a(r-1)}\right) C_r = \delta_{n1} I \qquad \text{for } 1 \leqq n \leqq N,$$

$$\sum_{r=1}^{N+1} I\, e^{-a(1-r)^2+b(1-r)} Q_0 \left(1 - e^{-2ar}\right) C_r = 0 \qquad \text{for } n = N+1,$$

which again provides us with a reduced system since $1 - e^{-2a(r-1)}$ vanishes for $r = 1$. We continue in this way and ultimately obtain

$$\sum_{r=N+1}^{N+1} I\, e^{-a(N+N+2-p-r)^2+b(N+N+2-p-r)} Q_0 \left(1-e^{-2ar}\right)\left(1-e^{-2a(r-1)}\right)\cdots\left(1-e^{-2a(r-N)}\right) C_r = \delta_{p,N+1} I \qquad \text{for } p = N+1,$$

which gives

$$C_{N+1}^{-1} = \prod_{p=1}^{N+1}\left(1 - e^{-2ap}\right) Q_0 .$$

Note that this is independent of b.

SUBSAMPLING PREDICTION

One question of some importance in prediction theory is to what extent the prediction error is affected if a subprocess of the given process is used instead of the original process itself. To be specific let us compare the linear prediction error for a given multivariate wide-sense stationary process $X_0, X_1, X_2, \cdots X_{N+1} = X_{(K+1)\beta}$ with the error for an equally spaced subprocess of this, $X_0, X_\beta, X_{2\beta}, X_{3\beta}, \cdots X_{(K+1)\beta}$, the prediction in each case to be of the type $X_{N+1} = X_{(K+1)\beta}$ on the basis of the preceding ones going back to X_0. According to the linear least squares theory the predicted values for X_{n+1} in the two cases are

$$\hat{X}_{N+1} = -\sum_{r=0}^{N} B_{N+1}^{-1} B_r X_r$$

and

$$\hat{X}_{(K+1)\beta} = -\sum_{s=0}^{K} \tilde{B}_{(K+1)\beta}\, \tilde{B}_{s\beta}\, X_{s\beta} .$$

In view of the wide-sense stationarity we have $\tilde{R}_{s\beta} = R_{s\beta}$ where $\tilde{R}_{s\beta}$ is the covariance matrix of the subprocess. The system of equations for $\tilde{C}_{s\beta} \equiv \overline{\tilde{B}^T_{s\beta}}\, \tilde{B}_{(K+1)\beta}$ is then

$$\begin{aligned}
&R_{(K+1)\beta}\tilde{C}_0 + R_{K\beta}\tilde{C}_\beta + R_{(K-1)\beta}\tilde{C}_{2\beta} + \cdots + R_0\tilde{C}_{(K+1)\beta} = I ,\\
&R_{K\beta}\tilde{C}_0 + R_{(K-1)\beta}\tilde{C}_\beta + R_{(K-2)\beta}\tilde{C}_{2\beta} + \cdots + \overline{R^T_\beta}\tilde{C}_{(K+1)\beta} = 0 ,\\
&R_{(K-1)\beta}\tilde{C}_0 + R_{(K-2)\beta}\tilde{C}_\beta + R_{(K-3)\beta}\tilde{C}_{2\beta} + \cdots + \overline{R^T_{2\beta}}\tilde{C}_{(K+1)\beta} = 0 ,\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
&R_0\tilde{C}_0 + \overline{R^T_\beta}\tilde{C}_\beta + \overline{R^T_{2\beta}}\tilde{C}_{2\beta} + \cdots + \overline{R^T_{(K+1)\beta}}\tilde{C}_{(K+1)\beta} = 0 .
\end{aligned}$$

It will be noted that this array is essentially the same as the original array with the exception that terms corresponding to the elements cast out in the subsampling process are suppressed.

It appears that a determination of the difference in the two prediction errors $C^{-1}_{N+1} - C^{-1}_{(K+1)\beta}$ for the general case is not feasible. In special cases certain information may be gained. For example. Equation 28 shows that for a quasi white noise process $C^{-1}_{N+1} - \tilde{C}^{-1}_{(K+1)\beta} = O(\epsilon^2)$, since the terms in ϵ^0 and ϵ involve only R_0. Also, the last two examples are simple enough that the error difference can be obtained explicitly. In the exponential decay process $R_{s\beta} = e^{-a|s\beta|}\, Q_0 = e^{-a\beta|s|}\, Q_0$ so all we need do is treat the subprocess the same as the original with a replaced by $a\beta$. Therefore, $C^{-1}_{N+1} - \tilde{C}_{(K+1)\beta} = \left(e^{-2a\beta} - e^{-2a}\right) Q_0$, and in the quadratic exponential decay process $R_{s\beta} = e^{-a\beta^2 s^2 + b\beta s}\, Q_0$, whence

$$C^{-1}_{N+1} - \tilde{C}^{-1}_{(K+1)\beta} = \left[\prod_{p=1}^{N+1}\left(1 - e^{-2ap}\right) - \prod_{r=1}^{K+1}\left(1 - e^{-2a\beta^2 r}\right)\right] Q_0 .$$

DIFFERENCING OF MULTIVARIATE PROCESSES

The first-order divided difference in the multivariate case is defined as

$$X[su] \equiv (s-u)^{-1}\left\{X_s - X_u\right\} ,$$

the second-order divided difference is defined by

$$X[tsu] \equiv (t-u)^{-1}\left\{X[ts] - X[su]\right\} = (t-u)^{-1}(t-s)^{-1}(s-u)^{-1}\left\{(s-u)X_t + (u-t)X_s + (t-s)X_u\right\}; \quad (29)$$

and so on. Just as in the case of differentiation the differencing of a polynomial gives a polynomial of lower degree. In particular, the second-order divided difference of a second-degree polynomial (with column matrices for coefficients) is a constant (matrix), and the third-order difference is zero.

Suppose Ξ_0, Ξ_1, Ξ_2, and Ξ_3 are random variables with Ξ_2 and Ξ_3 orthogonal, and let λ be a real number. Then the second-order divided differences of the process

$$X_t = \Xi_0 + \Xi_1 t + \Xi_2 t^2 + \Xi_3 e^{2\pi i\lambda t}$$

form a wide-sense stationary process; that is,

$$E\left\{X[t_1+h, s_1+h, u_1+h]\,\overline{X[t_2+h, s_2+h, u_2+h]^T}\right\}$$

is independent of h:

$$E\left\{X[t_1+h, s_1+h, u_1+h]\,\overline{x[t_2+h, s_2+h, u_2+h]^T}\right\}$$

$$= E\left\{\left(\Xi_2 + (t_1-u_1)^{-1}(t_1-s_1)^{-1}(s_1-u_1)^{-1}\left[(s_1-u_1)e^{2\pi i\lambda(t_1+h)} + (u_1-t_1)e^{2\pi i\lambda(s_1+h)} + (t_1-s_1)e^{2\pi i\lambda(u_1+h)}\right]\Xi_3\right)\left(\overline{\Xi_2^T}\right.\right.$$

$$\left.\left.+ (t_2-u_2)^{-1}(t_2-s_2)^{-1}(s_2-u_2)^{-1}\left[(s_2-u_2)e^{-2\pi i\lambda(t_2+h)} + (u_2-t_2)e^{-2\pi i\lambda(s_2+h)} + (t_2-s_2)e^{-2\pi i\lambda(u_2+h)}\right]\overline{\Xi_3^T}\right)\right\}$$

$$= E\{\Xi_2\overline{\Xi_2^T}\} + (t_1-u_1)^{-1}(t_1-s_1)^{-1}\cdots(s_2-u_2)^{-1}\left[(s_1-u_1)e^{2\pi i\lambda t_1} + (u_1-t_1)e^{2\pi i\lambda s_1}\right.$$

$$\left.+ (t_1-s_1)e^{2\pi i\lambda u_1}\right]\left[(s_2-u_2)e^{-2\pi i\lambda t_2} + (u_2-t_2)e^{-2\pi i\lambda s_2} + (t_2-s_2)e^{-2\pi i\lambda u_2}\right]E\{\Xi_3\overline{\Xi_3^T}\}$$

and this does not involve h.

This result extends immediately. The process

$$X_t = \sum_{k=0}^{p}\Xi_k t^k + \sum_{k=1}^{n}\Xi_{p+k}\,e^{2\pi i\lambda_k t}$$

for any set of real numbers, λ_k, and Ξ_0, Ξ_1, $\cdots$ Ξ_{p+n} random variables with Ξ_p, Ξ_{p+1}, $\cdots$ Ξ_{p+n} orthogonal has wide-sense stationary p^{th} order divided differences.

The algebraic polynomial part of X_t represents what is commonly called the "trend," while the trigonometric polynomial part is wide-sense stationary. In practice, the trend is commonly removed by fitting a polynomial to the data, say by least squares, leaving as residue the wide-sense stationary part. As the above work shows a wide-sense stationary process can also be obtained by forming the differences of sufficiently high order terms of the original process. Of course, the resulting wide-sense stationary process is not the same as the one obtained from removing the trend by curve-fitting methods. The former represents the differences of some order of the latter (to within the error in the curve fitting).

(Manuscript received January 21, 1966)

REFERENCES

1. Doob, J. L., "Stochastic Processes," New York: John Wiley and Sons, 1953.

2. Rosenblatt, M., "A Multi-dimensional Prediction Problem" *Arkiv für Matematik,* 3:407-424, 1958.

3. Helson, H., and Lowdenslager, D., "Prediction Theory of Fourier Series in Several Variables," *Acta Mathematica,* 99:165-202, 1958.

4. Masani, P., and Wiener, N., "The Prediction Theory of Multivariate Stochastic Processes I. The Regularity Condition," *Acta Mathematica,* 98:111-150, 1957.

5. Masani, P., and Wiener, N., "The Prediction Theory of Multivariate Stochastic Processes II. The Linear Predictor," *Acta Mathematica,* 99:93-137, 1958.

6. Grenander, U., and Rosenblatt, M., "Statistical Analysis of Stationary Time Series," New York: John Wiley and Sons, 1957.

NASA-Langley, 1966

"The aeronautical and space activities of the United States shall be conducted so as to contribute . . . to the expansion of human knowledge of phenomena in the atmosphere and space. The Administration shall provide for the widest practicable and appropriate dissemination of information concerning its activities and the results thereof."

—NATIONAL AERONAUTICS AND SPACE ACT OF 1958

NASA SCIENTIFIC AND TECHNICAL PUBLICATIONS

TECHNICAL REPORTS: Scientific and technical information considered important, complete, and a lasting contribution to existing knowledge.

TECHNICAL NOTES: Information less broad in scope but nevertheless of importance as a contribution to existing knowledge.

TECHNICAL MEMORANDUMS: Information receiving limited distribution because of preliminary data, security classification, or other reasons.

CONTRACTOR REPORTS: Technical information generated in connection with a NASA contract or grant and released under NASA auspices.

TECHNICAL TRANSLATIONS: Information published in a foreign language considered to merit NASA distribution in English.

TECHNICAL REPRINTS: Information derived from NASA activities and initially published in the form of journal articles.

SPECIAL PUBLICATIONS: Information derived from or of value to NASA activities but not necessarily reporting the results of individual NASA-programmed scientific efforts. Publications include conference proceedings, monographs, data compilations, handbooks, sourcebooks, and special bibliographies.

Details on the availability of these publications may be obtained from:

SCIENTIFIC AND TECHNICAL INFORMATION DIVISION

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

Washington, D.C. 20546